An Embedded AI system for Predicting and Correcting the Sensor-orientation of an Electronic Travel Aid during use by a Visually Impaired Person

Gagandeep Singh[2], Mohd. Nadir[3], Rachit Thukral[1], Varun Gambhir[1],

Piyush Chanana[1] and Rohan Paul[1,2,3]

[1]School of Information Technology,
[2]Yardi School of Artificial Intelligence, [3]Deparatment of Computer Science Engineering,
Indian Institute of Technology Delhi, India
{piyush.chanana@sit.iitd.ac.in, anz228562@sit.iitd.ac.in,
aiy227511@iitd.ac.in, ird600069@iitd.ac.in,
rohan@cse.iitd.ac.in}

Abstract. People with visual impairment frequently rely on Electronic Travel Aids for navigation, with sensor-equipped guide canes being a prevalent choice. These ETAs alert users of nearby obstacles using directive sensing technology, effectively covering the navigation corridor of the person motion, thus facilitating obstacle avoidance. However, the efficacy of these devices depends upon the user's ability to maintain them in the correct orientation, a requirement that may not always be met due to various factors, including lack of training and excess cognitive load on the user in mobility, resulting in injuries. This paper proposes the development of an advanced AI-based embedded system, designed to be integrated into a traditional white guide cane. This proposed system predicts the instantaneous orientation of the ETA from raw proprioceptive measurement of ETA's instantaneous velocity and acceleration. If the estimated angle lies beyond a nominal range, audio or vibration feedback is provided proportional to the error.

Keywords: Embedded System; Artificial Intelligence; Assistive Technologies.

1 Introduction

Electronic Travel Aids (ETAs) are instrumental in enhancing the mobility of persons with visual impairment (PVI), offering the potential for greater independence and safety. Many of these aids are integrated into canes, providing the ability to detect obstacles from the ground to head height using one or more sensors. These sensors alert users of obstacles in their path by emitting audio-vibratory feedback, significantly reducing the risk of collisions and potential injuries. Despite the apparent benefits of ETAs, their effective utilization depends on maintaining the proper alignment of sensors for obstacle detection as the user navigates their environment. Trainers play a crucial role in monitoring the PVI's use of the ETA, providing immediate feedback and guidance when sensors are tilted sideways or held in an incorrect orientation. Based on our discussions and feedback from over 100 ETA users in India from 2015-22, it was observed that the majority do not find the devices effective until they have received comprehensive mobility training and engaged in regular practice sessions. However, in areas where there is a scarcity of experienced mobility trainers, access to training for correct use and holding of ETA is unmet, resulting in incomplete and incorrect training by help of other untrained family members or by self, leading to a continued reliance on others for navigation, negating the potential benefits of ETAs for PVIs.

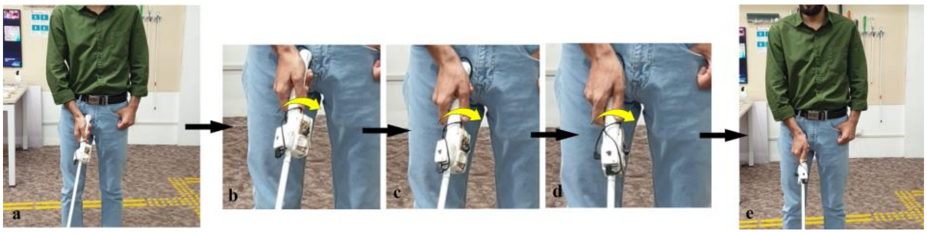

Fig. 1. Orientation correction by user. (a) is the Visually Impaired person holding the ETA incorrectly. (b), (c), (d) shows the ETA being transitioned to the correct orientation aided by our system's auditory and vibratory feedback. (e) shows the Visually impaired person correctly corrected the orientation of the ETA.

This work presents design, implementation and evaluation of a cane mountable AI based intelligent embedded system that learns the ETA holding style of the user, continuously detects the ETA orientation and guides the user to self-correct it through intuitive audio-vibratory feedback. Thereby, minimizing the dependence on the trainer and facilitating self-learning.

2 Related Works

In recent years, there have been several efforts to augment the white cane with sensing technology to improve its detection range. These advancements provide a pre-warning to users about obstacles in their navigation corridor. These efforts are based on ultrasonic ranging [1-3], lasers [4], wearable cameras [5,6] and sensors combined with a camera [7] to not only avoid obstacles but also identify them, enhancing the overall navigation experience for PVI. Also, there are devices that recognize user gestures [8]. Furthermore, existing research has addressed rule-based solutions to orientation correction by adjusting sensor angles to maintain a constant angle from the ground [9].

However, these solutions often fall short in accounting for incorrect device use, all the available solutions assume the PVI is holding the ETA in correct fashion. Hence, an unparalleled necessity arises for a comprehensive orientation detection and correction system that allows individuals to self-correct the orientation.

3 Technical Approach

We have developed a system which comprises of a battery powered Embedded processing device interfaced to an Inertial Measurement Unit (IMU) sensor, mounted on a cane. This system uses proprioceptive sensing to predict the orientation of the ETA as the PVI is holding it. We train an auto-regressive neural network that predicts the instantaneous orientation of the cane (continuous values along the vertical and rotational axes). The predictions are then discretized and (i) converted to modulated audio buzzer sounds and vibratory output to the user and (ii) visualized and recorded through a graphical display and logs on the computer for the mobility trainer to analyze.

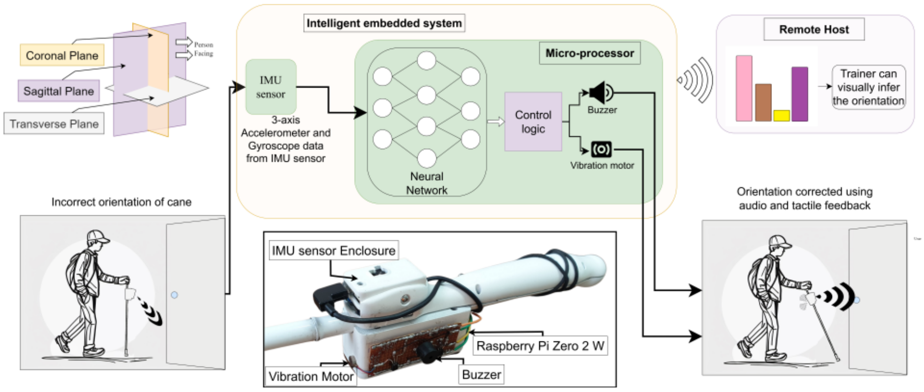

Fig. 2. Overall System Architecture. The system consists of an Embedded processing device with integrated wireless transmission unit and an IMU sensor, A neural network implemented on the device which takes raw sensor data and predicts the orientation of the device, these predictions are modulated for controlling auditory-sensory feedback and wirelessly transmitted to remote server for visual inference.

3.1 Neural Network and Training

Our aim is to design a neural network that takes sensor measurements as input and predicts the instantaneous angle in both sagittal and coronal plane. Since, the choice of the architecture was based on factors like performance, computational requirements, parameter size, power consumption and inference time when deployed on the system. Thus, we tested two different neural architectures: LSTM and Transformer based network. The Noisy raw data from the sensor was given to the neural network to predict the angle of the ETA. The angle predicted by the network is of 4 independent classes, each class $\in [0,1]$. The predicted classes are Slant, Right, Left and Vertical. We define $\hat{y}$ = Prediction($x_{0:t}$), where "$x_{0:t}$" is the raw sensor signal of sequence length 300 and dimensionality of 6. whereas the "prediction" is the neural network prediction.

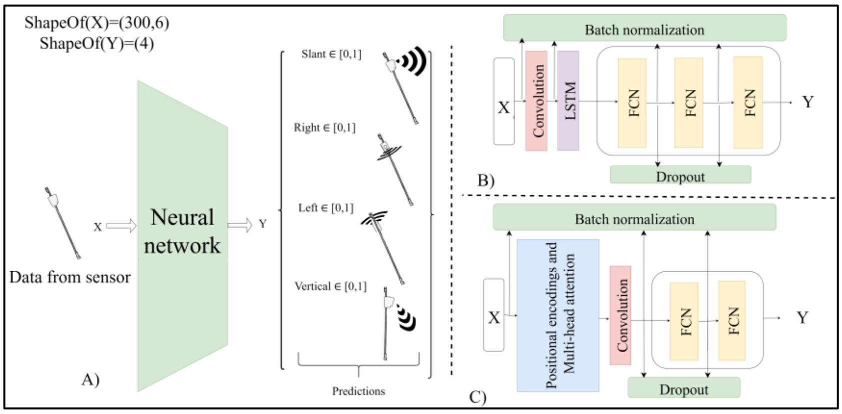

Fig. 3. Neural Network Architecture. A) The neural network takes input(X) from sensor on ETA and predicts(Y) orientation, B) Network architecture of LSTM based neural network contains 1391 learnable parameters, C) Network architecture of transformer based neural network contains 3412 learnable parameters.

The Neural Networks were optimized using TFlite. TFlite optimizes the models using operations like quantization and pruning. Additionally, it uses FlatBuffers for the optimization of the network. With FlatBuffers, the data can be accessed directly from the serialized buffer without parsing/unpacking. Allowing TFlite to load and run models faster and with less memory consumption.

Table 1. Neural Network Architecture and parameters of LSTM and Transformer models.

LSTM			Transformer		
Layer name	**Output shape**	**Parameters**	**Layer name**	**Output shape**	**Parameters**
Batch normalization	300,6,1	4	**Batch normalization**	300,6	24
Convolution	9,4,1	151	-	-	-
Batch normalization	9, 4, 1	4	**Multi-head Attention**	300,6	1424
LSTM	8	320	**Convolution**	9,1,4	1204
FCN	16	144	-	-	-
Batch normalization	16	64	**Batch normalization**	36	144
Dropout	16	0	**Dropout**	36	0
FCN	32	544	**FCN**	16	592
Batch normalization	32	128	**Batch normalization**	16	64
Dropout	32	0	**Dropout**	16	0
FCN	4	132	**FCN**	4	68

3.2 Data Collection and Training Data Generation

To create our training set, 3-axis acceleration and gyroscope data was recorded from an Inertial Measurement Unit (IMU) sensor at a sampling rate of 6660 Hz. Labelling raw IMU data was challenging due to high data rate, noise (due to ETA holding/tapping by the person) and high cost of expensive motion capture systems to record ground truth. Hence, to enable large scale data generation without requiring specialized external hardware, we used a combination of classical algorithms for pre-processing and scoring of the data to generate labels. In the data labeling process, labels were assigned continuous values to accurately indicate the angle of the ETA within the plane. We have labelled the data such that all the classes independently provide the angle of the ETA.

Data Collection. We developed a prototype system with a battery and a micro-controller with which we interfaced a microSD module such that we can collect the raw data from the sensor and store it in the storage device. Using this system, we collected the raw sensor data in 3 passes, (i) Slowly moving the cane from almost vertical position (around 80 degrees) to almost horizontal position (around 10 degrees) in about 2 minutes. This data was collected to generate labels for the slant class, we will refer to this data as “raw sensor slant data”. (ii) Slowly moved the cane from the position of extreme right to extreme left in around the same time, this data was collected to label the right and left orientation, we will refer to this data as “raw sensor right-left data”. (iii) For the collection of the vertical class in which the cane was held in vertical position for about 30 seconds and we will refer to this data as “raw sensor vertical data”. After the collection of raw data, we then recorded a reference signal which was a few seconds of cane held in correct orientation.

Training Data Generation. The data generation or labeling process uses a similarity measure technique, we used Dynamic Time Warping (DTW) [10] algorithm to find the distance/similarity of the unlabeled data and a single correct labelled data sample. The DTW is a useful technique for comparing time domain signals. For the similarity calculation of the data with a reference signal, we used the 3-accelerometer axis of both the unlabeled and reference signal, we omitted the gyroscope axis data from the signal to avoid the dependence of velocity of motion on the labelling of data.

In the process of data generation, (i) We filtered the raw sensor data using a series of time domain filers. (ii) We segmented the filtered unlabeled data from the filters into sequence lengths of 300. We used this unlabeled data and calculated the distance of this sequence and the reference signal to compute a distance between the unlabeled data and the reference signal using DTW algorithm, (iii) We normalized the DTW distance/similarity scores to make labels for the slant class. Similarly, we did the same for the right left orientation data using a different reference signal to get the labels. We define the pseudo labels generated by the label generating algorithm as $\tilde{y} = Estimate(x_{0:t})$, where “*Estimate*” is the Estimated labels generated by the supervised label generation process. We train the Neural Network using backpropagation of the loss. Mathematical equation of Loss = $\mathcal{L}(\tilde{y}, \hat{y})$ is as follows.

$$\mathcal{L}(\tilde{y}, \hat{y}) = \alpha\mathcal{L}_{slant}(\tilde{y}, \hat{y}) + \beta\mathcal{L}_{right}(\tilde{y}, \hat{y}) + \gamma\mathcal{L}_{left}(\tilde{y}, \hat{y}) + \delta\mathcal{L}_{vertical}(\tilde{y}, \hat{y}) \quad (1)$$

$$\mathcal{L}_{class}(\tilde{y}, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(\tilde{y}_{class} - \hat{y}_{class})^2 \quad (2)$$

The co-efficient, $\alpha=\beta=\gamma=\delta=1$. (3)

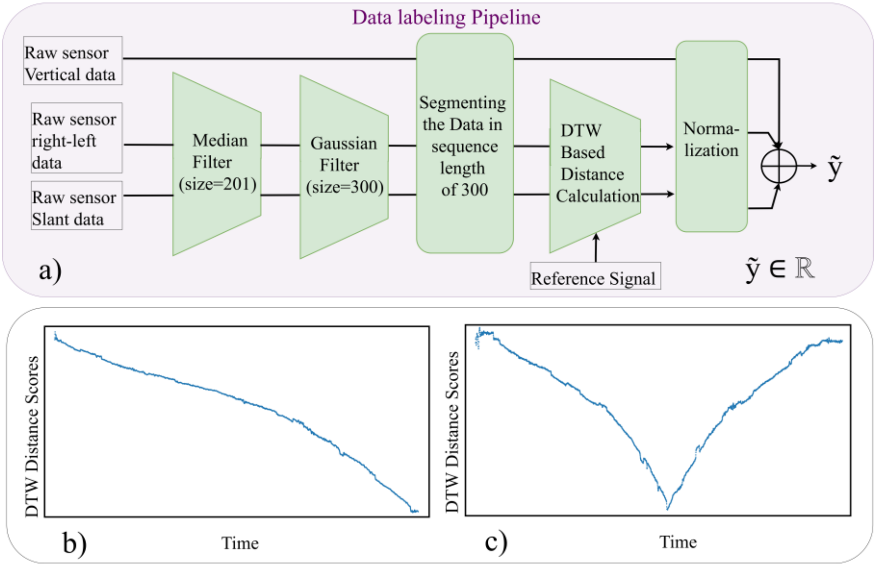

Fig. 4. Label Generating Process. (a) The collected raw sensor data for *slant*, and *right-left* is passed through a median and gaussian filter. After filtering the data is segmented to a sequence length of 300. The segmented data is passed to a Dynamic Time Warping based labelling algorithm for labelling. The resulting data labels are combined to provide the labels $\tilde{y}$. (b) Normalized DTW distance scores generated using the above pipeline for the slant data, as the ETA is moved from vertical position to horizontal position (near ref. signal) the distance score is decreasing. (c) DTW distance score for right-left data, as the ETA moved from extreme right position to correct position (near ref. signal) then continued to move towards extreme left position the distance scores first decreased then increased.

3.3 Hardware Implementation

Utilizing the TensorFlow Lite (TFlite) [11] API, we convert the neural models to low-footprint embedded processor compatible TFlite format and deployed them on the device for real time operation. The system calculated the neural prediction and generated the Pulse Width Modulation (PWM) signal using the same core of the CPU. To circumvent the delay due to PWM signal's time-period, a multi-core processing approach was employed. Specifically, one microprocessor core was allocated for neural network computations, while another core was dedicated to PWM generation.

Considering the constraints of limited RAM on the microprocessor, we implemented memory swapping with flash storage, resulting in the successful real-time operation of the system. For user feedback mechanisms, auditory and sensory cues were employed, utilizing a buzzer and vibration motor. These components were controlled by a variable duty cycle signal originating from the microprocessor. The system was intricately designed to alert right and left orientations using vibrator feedback, while slant and vertical orientations were alerted using buzzer beeps as feedback. All the system operations were powered by a 4.44-Watt hr. battery. Moreover, the predictions generated by the neural network were wirelessly transmitted to a remote server connected to the system, facilitating visual inference.

3.4 Graphical Interface and Logging

To effectively interpret the system's predictions, it was imperative to develop a Graphical User Interface (GUI) capable of visualizing the results in real-time. To achieve this, a web server implementation was employed, owing to its rapid data transmission capabilities, and reduced computational demands on the server side. The neural network's predictions are stored within the system, from which a host device retrieves the data and displays the graph, the data is also logged for an orientation and mobility (O&M) trainer to analyze.

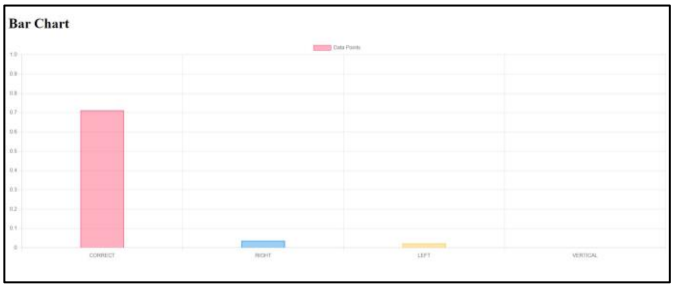

Fig. 5. Visual Inference. The Graphical User Interface on remote host shows the orientation of the ETA. The latency of the interface is minimized to visualize the ETA orientation in real-time.

4 Evaluation and Results

We evaluate the performance of our system for visually impaired individuals. The evaluation covers three primary areas: predictive accuracy of the LSTM and Transformer models, user adaptability in orientation correction, and the system's operational efficiency on low-power edge devices.

4.1 Model Accuracy

Our system underwent evaluation on both neural networks. The labelled dataset was partitioned into training, testing, and validation sets, consisting of 4013, 1255, and 1004 samples, respectively. The performance of each network was assessed on the test dataset, revealing an accuracy of 96.73% for the LSTM and 97.72% for the Transformer network. Although the labelled dataset consists of pseudo labels generated by DTW technique, in our findings they align closely with what would be expected from actual human-annotated data. Both networks achieve high accuracy, they still struggle in scenarios where the right, and left orientations are very close to correct orientation.

		Actual Class			
		Slant	Right	Left	Vertical
Predicted Class	Slant	429	13	0	0
	Right	17	263	0	0
	Left	8	0	245	0
	Vertical	3	0	0	277

(a)

		Actual Class			
		Slant	Right	Left	Vertical
Predicted Class	Slant	442	0	0	0
	Right	15	265	0	0
	Left	13	0	240	0
	Vertical	0	0	0	280

(a)

Fig. 6. Confusion matrix. The above image shows the confusion matrix of the (a) LSTM and (b) Transformer neural network. The confusion matrix shows the accuracy of the models on the pseudo labels generated by the labels generating algorithm. Both the models perform worse for the slant class compared to other classes.

4.2 Self-Correction by User

In the control trial, an adult visually impaired male participant was provided with the cane in arbitrary orientations and tasked with adjusting it to the correct orientation using feedback from the buzzer and vibration. Across 10 iterations with a PVI, the following observations were made: The user successfully corrected the cane eight times, with an average correction time of less than 5 seconds. The user also added that "*this device can be very useful for mobility training as they don't have to correct the first-time user again and again*". Notably, the user demonstrated an improvement in correction speed as the trial progressed. Recognizing the user-specific variability in slant preferences, we identified the need for fine-tuning the feedback mechanism to adapt to individual holding patterns. This adaptation can be achieved through modulation equation adjustments, eliminating the need for repetitive AI model training. The videos of user trial and GUI output can be seen here.

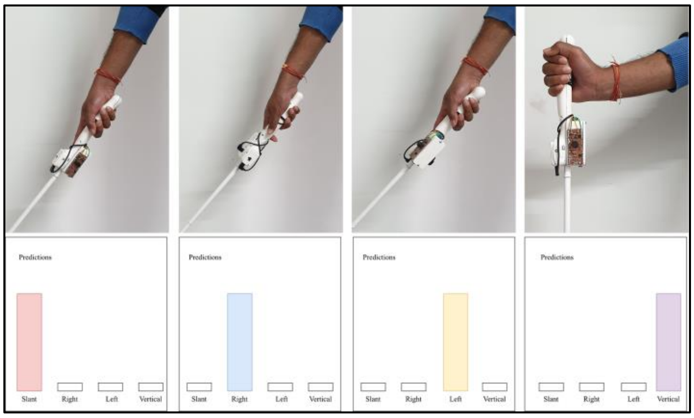

Fig. 7. ETA orientation and prediction. Views of different orientations of holding the cane with their corresponding orientation angle predictions on GUI. The continuous predictions are discretized and converted to buzzer sound and vibration feedback to the visually challenged user as well as logged/shown to the mobility trainer.

4.3 Real-Time Evaluation.

The two neural architectures had a small memory footprint with LSTM having 107 KB and the transformer having 42 KB of space in the computer. Successful inference runs were achieved on the Raspberry Pi Zero 2W in 18 milliseconds and 48 milliseconds for the LSTM and transformer networks, respectively. In user trials, the Transformer network outperformed the LSTM-based counterpart in prediction of correct state.

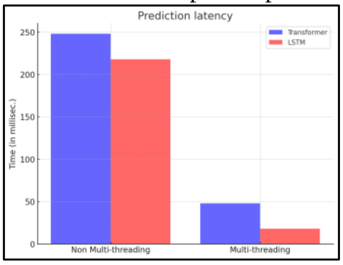

Fig. 8. Neural Network comparison. The prediction latency of the neural networks before and after multi-threading of neural network calculations and GUI calculations.

Additionally, a noteworthy observation revealed that the transformer architecture consumed approximately 1.5 times more power than its LSTM counterpart during inference. This substantial power consumption resulted in elevated temperatures within the microprocessor, highlighting a critical consideration for energy efficiency in system deployment. Furthermore, the graphical user interface (GUI) offers a real-time visualization of cane orientation for a trainer or family member to analyze, updating at a refresh rate of 20 milliseconds, enhancing user interaction.

5 Conclusions

In our research, we developed a system that detects and corrects cane orientation in real-time, offering auditory and vibratory feedback for proper alignment. Evaluation shows the system's effectiveness in slow movement scenarios, yet its accuracy in orientation prediction declines when rapid motion is experienced. Additionally, it cannot accurately predict orientations beyond 90 degrees in the coronal plane nor provide the cane's angle in the transverse plane. Future efforts will aim to improve the model's robustness, especially in fast-motion situations, and to extend predictions to include transverse plane angles during movement.

Acknowledgement. Authors acknowledge research funding support from DST, GoI under TiDE program and from ICMR, GoI under National Center for Assistive Health Technologies (NCAHT) project at IIT Delhi. We thank volunteers from Saksham Trust who contributed to user evaluation. We thank Prof. P. V. M. Rao for providing inspiration and mentoring for undertaking this project. We acknowledge Ms. Sunita Negi for administrative support and the IIT Delhi HPC team for the high-performance computing infrastructure used for this work.

References

1. Chanana P, Paul R, Balakrishnan M, Rao P. Assistive technology solutions for aiding travel of pedestrians with visual impairment. Journal of Rehabilitation and Assistive Technologies Engineering. 2017;4. doi:10.1177/2055668317725993
2. Hoyle, Brian, and Dean Waters. "Mobility at: The batcane (ultracane)." Assistive technology for visually impaired and blind people. London: Springer London, 2008. 209-229.
3. C. Gearhart, A. Herold, B. Self, C. Birdsong and L. Slivovsky, "Use of ultrasonic sensors in the development of an Electronic Travel Aid," 2009 IEEE Sensors Applications Symposium, New Orleans, LA, USA, 2009, pp. 275-280, doi: 10.1109/SAS.2009.4801815.
4. Farcy, René, et al. "Electronic travel aids and electronic orientation aids for blind people: Technical, rehabilitation and everyday life points of view." Conference & Workshop on Assistive Technologies for People with Vision & Hearing Impairments Technology for Inclusion. Vol. 12. 2006.
5. Kedia, Rajesh, et al. "MAVI: Mobility assistant for visually impaired with optional use of local and cloud resources." 2019 32nd International Conference on VLSI Design and 2019 18th International Conference on Embedded Systems (VLSID). IEEE, 2019.
6. L. Everding, L. Walger, V. S. Ghaderi and J. Conradt, "A mobility device for the blind with improved vertical resolution using dynamic vision sensors," 2016 IEEE 18th International Conference on e-Health Networking, Applications and Services (Healthcom), Munich, Germany, 2016, pp. 1-5, doi: 10.1109/HealthCom.2016.7749459.
7. S. Bhatlawande et al., "AI Based Handheld Electronic Travel Aid for Visually Impaired People," 2022 IEEE 7th International conference for Convergence in Technology (I2CT), Mumbai, India, 2022, pp. 1-5, doi: 10.1109/I2CT54291.2022.9823962.
8. Shishir G. Patil, Don Kurian Dennis, Chirag Pabbaraju, Nadeem Shaheer, Harsha Vardhan Simhadri, Vivek Seshadri, Manik Varma, and Prateek Jain. 2019. GesturePod: Enabling On-device Gesture-based Interaction for White Cane Users. In Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology (UIST '19). Association for Computing Machinery, New York, NY, USA, 403–415. https://doi.org/10.1145/3332165.3347881
9. Manu Suresh, Jagannadh Pariti, and Tae Oh. 2019. Dynamic Sensor Orientation Unit for the Intelligent Mobility Cane. In Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '19). Association for Computing Machinery, New York, NY, USA, 548–550. https://doi.org/10.1145/3308561.3354627
10. H. Sakoe and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition," in IEEE Transactions on Acoustics, Speech, and Signal Processing, vol. 26, no. 1, pp. 43-49, February 1978, doi: 10.1109/TASSP.1978.1163055.
11. TensorFlow Authors. *TensorFlow Lite*. GitHub, 2024, https://github.com/tensorflow/tensorflow/tree/master/tensorflow/lite.